DoDPI99-R-0001

A Comparison of Computer Programs Designed to Evaluate Psychophysiological Detection of Deception Examinations: Bakeoff 1

Andrew B. Dollins, Ph.D., Donald J. Krapohl, M.A.,
and Donnie W. Dutton, M.S.

October 1999

Department of Defense Polygraph Institute
Fort Jackson, South Carolina 29207

20000713 014

DTIC QUALITY INSPECTED 4

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1999	Final Report (May 1996-November 1999)

4. TITLE AND SUBTITLE
A comparison of Computer Programs Designed to Evaluate Psychophysiological Detection of Deception Examination: Bakeoff 1

5. FUNDING NUMBERS
DoDPI96-P-0006

6. AUTHOR(S)
Andrew B. Dollins, Ph.D.; Donald J. Krapohl; and Donnie Dutton

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Defense Polygraph Institute
7540 Pickens Avenue
Fort Jackson, South Carolina 29207

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of Defense Polygraph Institute
7540 Pickens Avenue
Fort Jackson, South Carolina 29207

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
DoDPI99-R-0001

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Public Release, distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

DOLLINS, A. B., KRAPOHL, D. J., & DUTTON, D. W. A comparison of computer programs designed to evaluate psychophysiological detection of deception examinations: Bakeoff 1. October 1999, Report No. DoDPI99-R-0001. Department of Defense Polygraph Institute, Ft. Jackson, SC 29207.--This preliminary study was completed to determine if accuracy differences exist among computer programs designed to render decisions using data collected during psychophysiological detection of deception (PDD) examinations. The vendors of AXCON version 1.2, Chart Analysis version 5.1-090-17-097, Computerized Polygraph System version 2.2, Identifi version 1.43, and PolyScore version 4.0 participated in the project. The data from 97 PDD examinations (56 deceptive and 41 nondeceptive), for which examinee veracity was known only to the investigators, were sent to software vendors with the request that a decision of deception indicated, no deception indicated, or no opinion (e.g., inconclusive) be returned for each examination--using the vendors' software. The proportion of correct, erroneous, and no opinion decisions rendered by the five evaluated computer programs ranged from .71 to .77, .07 to .10, and .13 to .21, respectively. When no opinion decisions were excluded, the proportion of correct decisions ranged from .88 to .91. There were no statistically significant differences among the frequency of correct decisions, erroneous decisions, and no opinion decisions rendered by the five computer programs. There was no single examination in the data set that all of the computer programs classified as no opinion. Finally, all of the computer programs, except the Computerized Polygraph System, erroneously classified more nondeceptive examinees as deceptive (i.e., false positive) than deceptive examinees as nondeceptive (i.e., false negative).

Key-words: PDD, lie detection, computer program, software, comparison, veracity, accuracy, Chart Analysis, CPS, Identifi, PolyScore.

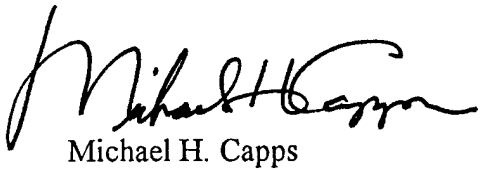
14. SUBJECT TERMS	15. NUMBER OF PAGES
	18
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLAS	UNCLAS	UNCLAS	

DTIC QUALITY INSPECTED 4

Director's Forward

In recent years, several computer programs have been designed to evaluate data collected during a psychophysiological detection of deception (PDD) examination. The accuracy of these computer programs and the potential utility for their use is of interest to the DoDPI. As part of the Congressional mandate to investigate mission essential elements of the PDD process the Institute compared the decision accuracy of five different computer programs. The psychophysiological data from 97 PDD examinations for which examinee veracity was known were sent to the four independent computer program developers with the instructions to provide the final decisions as determined by their software. The vendors were requested to limit their decision to deception indicated (DI), no deception indicated (NDI), or no opinion (inconclusive). The findings of the present study suggest that there were no statistically significant differences between the computer programs on the frequency of correct decision, erroneous decisions, and no opinion decisions. However, in this sample there was no single examination that all computer programs classified as no opinion (inconclusive) and many of the programs had higher false positives than false negatives in the final decision. Some methodological flaws within the study suggest that additional studies should be undertaken with better sampling controls. Moreover, the findings of the present study suggest much is needed to continue the development of reliable computer scoring programs.



Michael H. Capps
Director

This research was funded through the Department of Defense Polygraph Institute (DoDPI) project DoDPI96-P-0006, under contract number NXXX-XX-XX-I XXXX. The views expressed in this report are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

Acknowledgments

The authors wish to express thanks to the federal, state, and county agencies that supplied the psychophysiological detection of deception examinations which were evaluated, and to the software vendors who participated by evaluating those examinations. Special thanks to: N. Joan Blackwell; Robert P. Cole; William J. Gaschler, Jr.; Linda A. Landy; Jimmie W. Swinford; Rose M. Swinford; Earl D. Taylor, and Dan V. Weatherman of the Department of Defense Polygraph Institute for their contributions to this project. This project was funded by the Department of Defense Polygraph Institute as DoDPI96-P-0006. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

Abstract

DOLLINS, A. B., KRAPOHL, D. J., & DUTTON, D. W. A comparison of computer programs designed to evaluate psychophysiological detection of deception examinations: Bakeoff 1. October 1999, Report No. DoDPI99-R-0001. Department of Defense Polygraph Institute, Ft. Jackson, SC 29207.--This preliminary study was completed to determine if accuracy differences exist among computer programs designed to render decisions using data collected during psychophysiological detection of deception (PDD) examinations. The vendors of AXCON version 1.2, Chart Analysis version 5.1-090-17-097, Computerized Polygraph System version 2.2, Identifi version 1.43, and PolyScore version 4.0 participated in the project. The data from 97 PDD examinations (56 deceptive and 41 nondeceptive), for which examinee veracity was known only to the investigators, were sent to software vendors with the request that a decision of deception indicated, no deception indicated, or no opinion (e.g., inconclusive) be returned for each examination--using the vendors' software. The proportion of correct, erroneous, and no opinion decisions rendered by the five evaluated computer programs ranged from .71 to .77, .07 to .10, and .13 to .21, respectively. When no opinion decisions were excluded, the proportion of correct decisions ranged from .88 to .91. There were no statistically significant differences among the frequency of correct decisions, erroneous decisions, and no opinion decisions rendered by the five computer programs. There was no single examination in the data set that all of the computer programs classified as no opinion. Finally, all of the computer programs, except the Computerized Polygraph System, erroneously classified more nondeceptive examinees as deceptive (i.e., false positive) than deceptive examinees as nondeceptive (i.e., false negative).

Key-words: PDD, lie detection, computer program, software, comparison, veracity, accuracy, Chart Analysis, CPS, Identifi, PolyScore.

Table of Contents

One of the advantages of using a computerized polygraph to collect data during a psychophysiological detection of deception (PDD) examination is that computer software can be used to evaluate the physiological data. Today's examiners have several such programs which are designed to evaluate data collected during a PDD examination to choose from. The question then becomes one of determining which, if any, computer program provides the greatest accuracy when evaluating PDD examinations.

Method

This project was completed to compare the accuracy of computer programs designed to evaluate PDD examinations. In late 1997 the physiological data from 103 PDD examinations were selected from a database of confirmed cases maintained by the Department of Defense Polygraph Institute. All of the data were collected using an Axciton computerized polygraph system (Axciton Systems, Inc., Houston, TX). The Axciton polygraph records respiratory, cardiovascular (via an ascultatory cuff technique), and electrodermal activity. The manufacturer has not published details regarding the instrument, but basic specifications can be found in Cestaro (1997). Each examination consisted of at least three charts, as verified by a certified government examiner. The government examiner also classified the examinations as either Modified General Question Test (MGQT) or Zone Comparison Test (ZCT) format. No attempt was made to delineate among the format variations (e.g., single issue, multi-issue, and multi-facet; US Air Force MGQT vs. US Army MGQT; Backster ZCT vs. DoDPI ZCT; etc.). The examinations were classified as confirmed deceptive if the examinee signed a confession which included a statement describing the details of the crime, or if there was irrefutable corroborating evidence (e.g., positive urinalysis, fingerprints, possession of incident related paraphernalia, etc.) linking the examinee to the crime under investigation. Examinations were classified as confirmed nondeceptive if the examinee was cleared of involvement in the crime by confession of another person.

The data were sent to vendors of computer programs designed to evaluate PDD examinations. The vendors were asked to evaluate the examinations and return decisions of deception indicated (DI), no deception indicated (NDI), or no opinion (NO). None of the vendors had, to the best of our knowledge, seen the 103 examinations before this project and they were not told examinees' veracity until after all vendors had responded. The computer programs included AXCON version 1.2 and Chart Analysis version 5.1-090-17-097 by Axciton Systems, Inc. (Houston, TX); the Computerized Polygraph System (CPS) version 2.2 by the Stoelting Company (Wood Dale, IL); Identifi version 1.43 by Identifi (Olympia, WA); and, PolyScore version 4.0 by the Johns Hopkins University Applied Physics Laboratory (Laurel, MD). The Axciton AXCON program was under development by Axciton Systems, Inc. and was not available to the public when testing was done. The computer programs named Chart Analysis, AXCON, CPS, Identifi, and PolyScore are, to the best of our knowledge, the only software currently available to evaluate and render decisions using data from specific issue PDD examinations.

All of the computer programs tested were able to read the proprietary Axciton data format except the CPS system. The data sent for CPS evaluation were converted to text using a

program provided by Axciton Systems, Inc. The CPS analysis differed in that (a) the conversion program rounded question onset marks to the nearest full second (according to Bell, Raskin, Honts, & Kircher, 1999; CPS only scores electrodermal reactions if they begin at least .5 seconds after question onset) and (b) the CPS program was developed using a true measure of skin conductance (the Axciton system measures a hybrid of skin conductance and skin resistance). The influence of these differences is not known.

When reviewing the cases, in preparation for reporting results, we observed that the question content of some cases did not correspond to notations in our master database. We thus attempted to reconfirm each of the 103 examinations with the original testing agency. We were unable to confirm three cases because the testing agency had moved and the case files were lost. Three additional cases were discarded due to poor data quality as determined by a panel of DoDPI PDD examiners. The computer programs rendered either correct or NO decisions on each of these cases, so discarding them decreased, rather than increased, the accuracy of decisions reported here.

Of the 97 examinations analyzed, 64 (28 deceptive and 36 nondeceptive) were collected by federal agencies (e.g., US Army Criminal Investigations Division and US Postal Service). The remaining 33 examinations (28 deceptive and 5 nondeceptive) were collected by nonfederal agencies (e.g., Birmingham Police Department, Birmingham, AL; Clayton County Sheriff's Office, Jonesboro, GA; Marion County Sheriff's Office, Ocala, FL; Mobile County Sheriff's Office, Mobile, AL; South Carolina Law Enforcement Division, Columbia, SC).

The reported sample was composed of 44 MGQT examinations (27 deceptive and 17 nondeceptive) and 53 ZCT examinations (29 deceptive and 24 nondeceptive). Eighty-five examinations were of suspects under investigation, eight were of witnesses, and four were of victims. Twenty eight of the examinees were female and 69 were male. Of the 70 examinees for which race was available, 25 were African American, 39 were Caucasian, four were Hispanic, one was a Native American, and one was of South Pacific heritage. The examinee age range, for the 94 examinees for which age data were available, was 14 to 70 years with a mean of 32.6 ($\underline{SD}$ = 10.83).

Results

The frequency of decisions rendered by each computer program is presented in Table 1. Results of analyses using Cochran's Q statistic (Siegel & Castellan, 1988), which is analogous to a Repeated Measures Analysis of Variance for nonparametric data, indicated that there were no statistically significant differences among the proportion of correct ($\underline{Q}[4] = 3.243$, $\underline{p} = .518$), incorrect ($\underline{Q}[4] = 2.061$, $\underline{p} = .724$), and NO ($\underline{Q}[4] = 3.804$, $\underline{p} = .433$) decisions rendered by the computer programs. The frequency data are presented as proportions in Table 2. The Cochran's Q analyses were only calculated with NO decisions included because it is not, to the best of our knowledge, possible to calculate a repeated measures analysis unless all groups have an equal number of observations. The proportion of correctly identified deceptive examinees was between .732 and .893 (N=56); if NO decisions were excluded the range was .911 to .980. The proportion of correctly identified nondeceptive examinees was between .537 and .683 (N=41); if NO

decisions were excluded, the range was .727 to .903. The overall proportion of correct decisions regarding examinee veracity was between .711 and .773 (N=97); if NO decisions were excluded, the range was .881 to .908. The proportion of NO decisions was between .134 and .216 of the 97 decisions rendered. While the number of observations are too few for meaningful analysis, breakout decision frequency tables for test format and testing organization are provided in Appendix A to assist readers in understanding the data. The raw data are provided in Appendix B.

Table 1
Frequency of Decisions by Subject Veracity

	Deceptive (n = 56)			Nondeceptive (n = 41)		
Computer Program	Correct	Incorrect	No Opinion	Correct	Incorrect	No Opinion
AXCON	50	1	5	24	9	8
Chart Analysis	49	2	5	22	8	11
CPS	41	4	11	28	3	10
Identifi	49	1	6	22	8	11
PolyScore	49	1	6	26	7	8

Table 2
Proportions of Correct Decisions (with SEMs)

	NO Decisions Included			NO Decisions Excluded			Proportion
Computer Program	DI	NDI	TOTAL	DI	NDI	TOTAL	NO Decisions
AXCON	.893	.585	.763	.980	.727	.881	.134
	(.041)	(.077)	(.043)	(.020)	(.078)	(.056)	(.035)
Chart Analysis	.875	.537	.732	.961	.733	.877	.165
	(.044)	(.078)	(.045)	(.027)	(.081)	(.060)	(.038)
CPS	.732	.683	.711	.911	.903	.908	.216
	(.059)	(.073)	(.046)	(.041)	(.053)	(.052)	(.042)
Identifi	.875	.537	.732	.980	.733	.888	.175
	(.044)	(.078)	(.045)	(.020)	(.081)	(.058)	(.039)
PolyScore	.875	.634	.773	.980	.788	.904	.144
	(.044)	(.075)	(.043)	(.020)	(.071)	(.051)	(.036)

Note. NO = no opinion.

A test for the significance of proportion differences (Bruning & Kintz, 1987) indicated that the proportion of correctly identified deceptive examinees was significantly greater than the

proportion of correctly identified nondeceptive examinees ($p < .005$) for the AXCON, Chart Analysis, Identifi, and PolyScore computer programs. This was true when NO decisions were included and excluded. There were no significant differences between the proportion of deceptive and nondeceptive examinees correctly identified by the CPS computer program.

The proportion of agreement between pairs of computer programs are presented in Table 3. Two evaluation systems were considered to be in agreement if both classified the same examination as DI, NDI, or NO. The proportion of decision agreement between evaluation systems ranged from .722 to .907. Most of the disagreements between pairs of evaluation systems were combinations including NO decisions (i.e., DI and NO or NDI and NO).

Table 3
Proportion of Agreement Between Pairs of Scoring Systems (n = 97)

Computer Program	Chart Analysis	CPS	Identifi	PolyScore
AXCON	.907	.753	.804	.856
Chart Analysis		.742	.784	.804
CPS			.722	.753
Identifi				.722

Table 4
Frequency of Disagreement Between Pairs of Scoring Systems and Between Pairs of Scoring Systems and Ground Truth (n = 97)

Computer Program	Chart Analysis	CPS	Identifi	PolyScore	Ground Truth
AXCON	0	0	1	1	10
Chart Analysis		0	0	1	10
CPS			1	1	7
Identifi				0	9
PolyScore					8

Note. A disagreement is defined as one scoring system classifying an examination as DI while another classifies the same examination as NDI.

The decisions made by the five computer programs are summarized as follows. All five computer programs agreed and correctly classified the veracity of 52 examinees (i.e., 36 deceptive and 16 nondeceptive). All programs agreed and incorrectly classified three nondeceptive examinees as deceptive. Of the remaining 42 examinations: 16 deceptive examinees were classified as DI or NO, 3 deceptive examinees were classified as NDI or NO, 16 nondeceptive examinees were classified as NDI or NO, and three nondeceptive examinees were

classified as DI or NO. Four examinees (i.e., 1 deceptive and 3 nondeceptive) received opposite classifications by the computer programs. That is, one or more computer programs classified the examinee as DI while other computer programs classified the same examinee as NDI. The disagreement frequencies are shown in Table 4.

Discussion

Three important conclusions may be drawn from this preliminary study. First, as far as we were able to determine from this sample, there are no statistically significant accuracy differences among the five computer programs evaluated. A second, less apparent, conclusion is that there was no single examination in the data set that all of the computer programs classified as NO. Finally, as may be seen in Tables 1 and 2, all of the computer programs, except CPS, erroneously classified more nondeceptive examinees as deceptive (i.e., false positive) than deceptive examinees as nondeceptive (i.e., false negative).

The most obvious flaw with this report is bias due to sampling error. The data were those sent to the Institute between July and October 1997. Although a few cases that were erroneously decided by the original examiner, the majority of the cases were correctly decided—which could have biased the sample. Contributors to the data base may have been reluctant to send cases they had missed. All of the cases were confirmed via confession of the examinee or another—which could have further biased the sample because the number of unconfirmed and uncomfirmable examinations is not known. No attempt was made to counterbalance, randomly assign, or otherwise control for bias due to examiner ability, originating agency rules, test format, examinee status (e.g. suspect, witness, victim), examinee or examiner gender, examinee or examiner race, or examinee age. In fact, the only things which can be said about this sample with any degree of certainty are that the sample is composed of actual field data and examinee veracity was not available to the vendors.

There were no experimental controls regarding the software used in testing. Vendors were allowed to test the data and report results to the best of their ability. Vendors could have performed expert editing, screening, or other manipulations that might not be available to the field examiner. As previously mentioned, all of the data were collected using an Axciton computerized polygraph. If a laboratory grade instrument had been used to collect the test data, the results may have been quite different. (A laboratory grade instrument would allow absolute, rather than relative, measurement of reactions. A laboratory grade instrument would also measure true skin conductance, rather than a hybrid of conductance and resistance.) Differences between the computer programs used in this test were not addressed. Assessments regarding operator training and skill using each computer program would be appropriate topics for future evaluations. In addition, we were unable to control for or assess the accuracy of features or criteria used by the computer programs. Only the Stoelting Company has published the decision criteria used by their software (Kircher & Raskin, 1988). Similar publications by other vendors would allow us to assess the predictive ability of features used by the different computer programs.

The implication of the first conclusion, that there are no statistically significant differences in the tested computer programs ability to predict examinee veracity, is self explanatory. The differences in decision accuracy observed in this sample could be due to sampling error—and not to the ability of one or more computer programs to predict examinee veracity. The implications of the second conclusion, that all of the computer programs did not agree on a single NO decision are not so self evident. It is sometimes difficult to understand the significance of something that doesn't exist. There were cases where all of the scoring systems made correct and incorrect decisions. There were a few cases where the scoring systems made diametrically opposed decisions. There was, however, no single case that all scoring systems labeled unscorable. It is difficult to justify the continued supposition that a NO decision should be omitted or labeled as "correct" when there is so little agreement regarding when a NO decision should be made. Perhaps the PDD discipline should expend some effort to clearly define the parameters which make a NO decision appropriate.

The third conclusion drawn from this data, that all scoring systems, except the CPS system, have a bias toward predicting that subjects are deceptive was unexpected. The bias could have developed because the algorithm training sets were similar for all except the CPS system. (Chart Analysis and AXCON, Identifi, and PolyScore were developed using at least some confirmed case data supplied by the Department of Defense Polygraph Institute. All of the confirmed case data was collected using an Axciton computerized polygraph. The CPS system was developed using data collected with a Stoelting CPS by the US Secret Service.) The bias could also be an artifact of the data set used in this project. Vendors should be aware of this potential bias–which should be examined in future studies.

In summary, this preliminary study may be flawed due to the failure to manipulate, randomize, or otherwise control for (a) the data sample, (b) the data quality, and (c) operator skill. The field samples do, however, provide an index of how software designed to evaluate PDD examinations and predict subject veracity will perform. The sample size ($\underline{n} = 97$) provided the statistical power to detect a 10% difference between scoring systems, had such a difference actually existed, with a probability of .80. We therefore present the conclusion that there are no statistically significant differences in the tested computer programs ability to correctly predict deception. We suggest that efforts be made to clearly define the parameters necessary for a decision of NO and that future studies assess the possibility that a false positive bias exists among computer programs designed to evaluate PDD examinations.

References

Bell, B. G., Raskin, D. C., Honts, C. R., & Kircher, J. C. (1999). The Utah numerical scoring system. Polygraph, 28, 1-9.

Bruning, J. L., & Kintz, B. L. (1987). Computational handbook of statistics (3rd. ed.). Glenview, IL: HarperCollins.

Cestaro, V. L. (1998). Instrumentation for presenting a known standard signal to the electrodermal activity channel for assessing response characteristics. Polygraph, 27(3), 188-209)

Kircher, J. D., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. Journal of Applied Psychology, 73, 291-302.

Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences (2nd. ed.). New York, NY: McGraw-Hill.

Appendix A

Subgroup Decision Frequencies

Decision Frequencies for Federal and Nonfederal Examinations

Computer Program	Deceptive			Nondeceptive		
	Correct	Incorrect	No Opinion	Correct	Incorrect	No Opinion
			Federal Examinations			
AXCON	25	1	2	23	7	6
Chart Analysis	25	1	2	21	6	9
CPS	20	2	6	26	2	8
Identifi	25	0	3	21	6	9
PolyScore	26	1	1	25	4	7
			Nonfederal Examinations			
AXCON	25	0	3	1	2	2
Chart Analysis	24	1	3	1	2	2
CPS	21	2	5	2	1	2
Identifi	24	1	3	1	2	2
PolyScore	23	0	5	1	3	1

Decision Frequencies for Zone Comparison Test and Modified General Question Test Examination Formats

Computer Program	Deceptive			Nondeceptive		
	Correct	Incorrect	No Opinion	Correct	Incorrect	No Opinion
			Modified General Question Test			
AXCON	25	1	1	8	5	4
Chart Analysis	24	1	2	8	4	5
CPS	21	2	4	9	1	7
Identifi	24	0	3	9	5	3
PolyScore	25	1	1	9	2	6
			Zone Comparison Test			
AXCON	25	0	4	16	4	4
Chart Analysis	25	1	3	14	4	6
CPS	20	2	7	19	2	3
Identifi	25	1	3	13	3	8
PolyScore	24	0	5	17	5	2

Appendix B

Raw Data
(A label key is located at the end of the table)

S#	Trth	Axcon	ChtAn	CPS	Idntf	PlySc	Format	Agncy	Sex	Stus	Race
1	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Cauc
2	DI	DI	DI	DI	DI	DI	MGQT	Fed	M	Spt	Cauc
3	DI	DI	DI	NO	DI	DI	ZCT	Fed	F	Spt	Afri
4	NDI	NO	NO	NDI	NDI	NDI	MGQT	Fed	F	Spt	Cauc
5	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	Hisp
6	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	
7	NDI	NO	NDI	NDI	NDI	NO	MGQT	Fed	M	Spt	
8	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	
9	NDI	NDI	NDI	NO	NO	NDI	MGQT	Fed	M	Spt	
10	NDI	DI	DI	NO	DI	DI	MGQT	Fed	M	Spt	
11	NDI	DI	DI	DI	DI	DI	ZCT	Fed	M	Spt	
12	NDI	DI	DI	NO	DI	DI	ZCT	Fed	M	Spt	
13	NDI	NDI	NDI	NDI	NDI	NDI	MGQT	Fed	M	Spt	
14	NDI	NDI	NDI	NDI	NO	NDI	ZCT	Fed	M	Spt	
15	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	
16	NDI	NDI	NO	NDI	NO	NDI	ZCT	Fed	M	Spt	
17	NDI	NDI	NO	NDI	NO	NDI	ZCT	Fed	M	Spt	
18	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Cauc
19	NDI	NO	NO	NDI	NDI	NO	ZCT	Fed	M	Spt	
20	DI	DI	DI	DI	DI	DI	MGQT	Fed	M	Spt	Hisp
21	NDI	DI	NO	NO	DI	NO	MGQT	Fed	M	Spt	
22	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	
23	NDI	NO	NO	NO	NO	DI	ZCT	NoFed	M	Spt	Cauc
24	NDI	NDI	NDI	NDI	NDI	NDI	MGQT	Fed	M	Spt	
25	DI	DI	DI	NO	DI	DI	MGQT	Fed	F	Spt	Afri
26	NDI	DI	DI	DI	DI	DI	MGQT	Fed	M	Spt	
27	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	
28	DI	DI	DI	DI	DI	DI	MGQT	NoFed	M	Spt	Cauc
29	DI	DI	DI	DI	DI	DI	MGQT	NoFed	F	Spt	Afri
30	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Cauc
31	DI	DI	DI	DI	DI	DI	MGQT	NoFed	M	Spt	Afri
32	NDI	NDI	NO	NDI	NDI	NDI	MGQT	Fed	M	Spt	
33	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	
34	NDI	NDI	NDI	NDI	NDI	NDI	MGQT	Fed	M	Spt	
35	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Cauc
36	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Afri
37	NDI	NO	NO	NDI	NO	DI	ZCT	NoFed	F	Spt	Cauc

S#	Trth	Axcon	ChtAn	CPS	Idntf	PlySc	Format	Agncy	Sex	Stus	Race
38	DI	DI	DI	NO	DI	DI	MGQT	NoFed	M	Spt	Cauc
39	DI	DI	DI	DI	DI	DI	MGQT	NoFed	M	Spt	Afri
40	NDI	NO	NO	NDI	NO	NDI	ZCT	Fed	M	Spt	
41	DI	DI	NO	DI	DI	DI	MGQT	NoFed	M	Spt	Afri
42	DI	DI	DI	DI	DI	DI	ZCT	Fed	F	Spt	Cauc
43	NDI	NDI	NDI	NDI	NDI	NO	ZCT	Fed	M	Spt	
44	DI	DI	DI	DI	DI	DI	ZCT	Fed	F	Spt	Afri
45	DI	DI	DI	NO	DI	DI	MGQT	Fed	F	Wtn	Afri
46	DI	NO	NO	NDI	DI	NO	MGQT	Fed	F	Spt	Cauc
47	DI	DI	DI	DI	DI	DI	MGQT	Fed	F	Spt	Cauc
48	NDI	NDI	NDI	NDI	NDI	NDI	MGQT	Fed	F	Spt	Cauc
49	DI	DI	DI	DI	DI	DI	MGQT	Fed	M	Spt	Hisp
50	NDI	NO	DI	NO	NO	NO	MGQT	Fed	F	Spt	Cauc
51	DI	DI	DI	DI	DI	DI	MGQT	Fed	M	Spt	Afri
52	DI	DI	DI	DI	DI	DI	MGQT	Fed	F	Wtn	Cauc
53	DI	DI	DI	DI	DI	DI	MGQT	Fed	M	Wtn	
54	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	
55	DI	DI	DI	NO	NO	DI	MGQT	Fed	M	Wtn	Cauc
56	NDI	NDI	NDI	NDI	NDI	NDI	MGQT	Fed	F	Spt	Afri
57	DI	DI	DI	DI	DI	DI	MGQT	Fed	F	Spt	Cauc
58	DI	DI	DI	NO	DI	DI	ZCT	Fed	F	Spt	Cauc
59	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Spt	
60	DI	NDI	NDI	NDI	NO	NDI	MGQT	Fed	M	Spt	Hisp
61	NDI	NDI	NDI	NDI	NO	NDI	ZCT	Fed	M	Spt	
62	DI	DI	DI	DI	DI	DI	ZCT	Fed	M	Spt	Cauc
63	DI	DI	DI	DI	DI	DI	ZCT	NoFed	F	Spt	Cauc
64	DI	DI	DI	DI	DI	DI	MGQT	Fed	M	Spt	Afri
65	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Cauc
66	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	NoFed	M	Vtm	Afri
67	DI	DI	DI	NO	DI	NO	ZCT	NoFed	M	Spt	Cauc
68	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Afri
69	DI	DI	DI	DI	DI	NO	ZCT	NoFed	F	Vtm	Cauc
70	DI	DI	DI	NO	DI	DI	ZCT	NoFed	M	Spt	Afri
71	DI	DI	DI	NO	DI	DI	ZCT	NoFed	M	Spt	Cauc
72	DI	DI	DI	DI	DI	DI	MGQT	NoFed	M	Spt	Cauc
73	DI	DI	DI	DI	NO	DI	ZCT	NoFed	F	Spt	Cauc
74	NDI	DI	DI	NO	DI	NO	MGQT	NoFed	M	Spt	Afri
75	DI	NO	NDI	NDI	NO	NO	ZCT	NoFed	M	Spt	Afri
76	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Cauc
77	NDI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Cauc
78	DI	DI	DI	DI	DI	DI	MGQT	NoFed	M	Wtn	Cauc

S#	Trth	Axcon	ChtAn	CPS	Idntf	PlySc	Format	Agncy	Sex	Stus	Race
79	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Afri
80	DI	NO	NO	NDI	NDI	NO	ZCT	NoFed	M	Spt	Afri
81	DI	NO	NO	NO	NO	DI	ZCT	NoFed	F	Spt	Afri
82	DI	DI	DI	DI	DI	NO	ZCT	NoFed	M	Spt	Afri
83	DI	DI	DI	DI	DI	DI	ZCT	Fed	M	Spt	Cauc
84	NDI	DI	DI	NO	NO	NDI	ZCT	Fed	F	Vtm	Cauc
85	DI	DI	DI	DI	DI	DI	ZCT	Fed	F	Spt	Cauc
86	DI	DI	DI	DI	NO	DI	MGQT	Fed	M	Spt	Asia
87	NDI	NDI	NDI	NDI	NO	NDI	MGQT	Fed	F	Wtn	Afri
88	NDI	NO	NO	NO	DI	NO	MGQT	Fed	F	Wtn	Afri
89	DI	DI	DI	DI	DI	DI	ZCT	Fed	M	Spt	Afri
90	NDI	NDI	NDI	NDI	NDI	NDI	ZCT	Fed	M	Wtn	Cauc
91	DI	NO	NO	NO	DI	DI	ZCT	Fed	F	Spt	Cauc
92	DI	DI	DI	DI	DI	DI	MGQT	Fed	M	Spt	Cauc
93	NDI	DI	NO	NO	NDI	NO	MGQT	Fed	M	Vtm	Cauc
94	DI	DI	DI	DI	DI	DI	ZCT	Fed	F	Spt	Cauc
95	DI	DI	DI	DI	DI	DI	MGQT	Fed	M	Spt	Afri
96	DI	DI	DI	DI	DI	DI	MGQT	Fed	F	Spt	Nati
97	DI	DI	DI	DI	DI	DI	MGQT	NoFed	F	Spt	Cauc

Omitted Cases

S#	Trth	Axcon	ChtAn	CPS	Idntf	PlySc	Format	Agncy	Sex	Stus	Race
1	DI	DI	DI	NO	DI	DI	MGQT	NoFed	F	Spt	Cauc
2	DI	DI	DI	DI	DI	DI	ZCT	NoFed	M	Spt	Afri
3	NDI	NDI	NDI	NDI	NDI	NDI	MGQT	Fed	M	Wtn	Cauc
4	NDI	NDI	NO	NDI	NDI	NDI	MGQT	Fed	M	Wtn	Cauc
5	NDI	NDI	NO	NDI	NDI	NDI	MGQT	Fed	M	Spt	Hisp
6	NDI	NDI	NO	NDI	NDI	NDI	MGQT	NoFed	M	Spt	Cauc

Appendix B Label Key
S# - Subject number
Trth - Ground truth
Axcon - AXCON ver 1.2 (Axciton)
ChtAn - Chart Analysis 5.1 (Axciton)
CPS - Computer Polygraph System 2.2 (Stoelting)
Idntf - Identifi 1.43 (Identifi)
PlySc - PolyScore 4.0 (John Hopkins Univ. Applied Physics Laboratory)
Format
ZCT - Zone Comparison Test MGQT - Modified General Question Test
Agncy - Agency
Fed - Federal
NoFed - Nonfederal
Sex
M - Male
F - Female
Stus - Status
Spt - Suspect
Wtn - Witness
Vtm - Victim
Race
Afri - African American
Cauc - Caucasian
Hisp - Hispanic
Asia - South Pacific heritage